DOCUMENT RESUME

ED 399 295 TM 025 610

AUTHOR Johanson, George A.
TITLE Absolute and Relative Grading Standards.
PUB DATE Apr 93
NOTE 10p.; Paper presented at the Annual Meeting of the American Educational Research Association (Atlanta, GA, April 12-16, 1993).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Achievement; *Grades (Scholastic); *Grading; *Performance Based Assessment; *Standards; Student Evaluation; *Test Content; Test Interpretation
IDENTIFIERS *Absolute Values; Compromise; *Relativism

ABSTRACT

K. K. Waltman and D. A. Frisbie (1994) observed that teachers and parents often interpret grades given to students in both absolute and relative senses. They conclude that this sort of interpretation is illogical and may indicate misunderstandings in several areas. Absolute and relative methods of assigning letter grades are approached from several perspectives in this paper, and the strengths and weaknesses of both approaches are identified and discussed. To use relative grading methods properly, a teacher needs a larger reference group than is usually available in typical classrooms. The appropriate use of absolute standards for a classroom test requires a clearly defined content domain and justified standards of performance. Appropriate use of either absolute or relative standards in practice is extremely difficult or limiting. It is argued that the compromise approach to grading, employing both absolute and relative considerations, that is often used in practice, has a number of desirable characteristics and may be a more reasoned alternative to grading than strict adherence to either an absolute or relative method. Reconciling the insights provided by judgments about test content and by judgments about groups yields realistic evaluations of student achievement. (Contains 27 references.) (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
GEORGE JOHNSON
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 399 295

ABSOLUTE AND RELATIVE GRADING STANDARDS

In a survey of parents and teachers, Waltman & Frisbie (1994) found that both absolute (criterion-referenced) and relative (norm-referenced) interpretations of a mathematics grade were often adopted by parents and teachers. Waltman & Frisbie understood this to mean that there was 'difficulty differentiating' between absolute and relative interpretations of achievement and that this may have been due to misunderstandings of various sorts. The purpose of this article is to review the relevant issues involved in standard-setting for purposes of classroom grading and to note that current practice may not be as illogical as it appears to be at first glance. Indeed, the current practice of using joint or compromise standards may be a quite rational alternative to total reliance on one (absolute or relative) standard alone.

The fifth standard of Standards for Teacher Competence in Educational Assessment of Students (AFT, NCME, & NEA, 1990) states: 'Teachers should be skilled in developing valid pupil grading procedures which use pupil assessments.' In the discussion immediately following this standard, the implication that valid grading procedures are known to the measurement community is made clear: 'The principles for using assessments to obtain valid grades are known and teachers should employ them.' Central among these principles is that of standard-setting.

With respect to traditional (non-contractual, letter grades A-F) methods of grading, there are basically three recommended ways of assigning letter grades to a set of raw scores from an objective classroom test to be found in the literature. These are as an absolute measure of attainment (a criterion- or domain-referenced grade), a relative measure of attainment (a norm-referenced grade), or as a compromise taking both absolute and relative standards into consideration.

While some measurement professionals take a relatively neutral view of this debate (e.g., Gronlund & Linn, 1990; Cangelosi, 1990; Sax, 1989; Airasian, 1991), others take positions with more certainty and make more or less definitive recommendations for practice. Those who find a relative standards approach more acceptable make statements such as: 'Our measurement technology is inadequate to provide grading on a meaningful absolute standard. The most meaningful standard is the normative performance of similar previous students.' (Hopkins, Stanley, & Hopkins, 1990, p. 329); 'Letter grades are invariably norm referenced.' (Oosterhof, 1990, p. 424). 'In traditional classroom situations, marks should be based on a normative interpretation.' (Mehrens & Lehmann, 1984,p. 522-3).

Those who advocate an absolute standards approach may make rather conservative statements such as: 'No solution is perfect, but it does seem that grading on the basis of preset fixed standards may promote effective communication about academic accomplishments more than any of the other procedures.' (Hills, 1981, p. 299), or 'In our opinion, comparisons with established

TM025610

BEST COPY AVAILABLE

ABSOLUTE AND RELATIVE GRADING STANDARDS

George A. Johanson
College of Education, Ohio University

A paper presented at the annual meeting of the
American Educational Research Association
Atlanta, GA
April 1993

ABSTRACT

In a recent article, Waltman & Frisbie observe that teachers and parents often interpret grades given to students in both an absolute and relative sense. They conclude that this sort of interpretation is illogical and may indicate misunderstandings in several areas. In this paper, absolute and relative methods of assigning letter grades are examined from several perspectives. The strengths and weaknesses of both approaches are identified and discussed. An argument is put forth that the compromise approach to grading, employing both absolute and relative considerations and often used in practice, has a number of desirable characteristics and may well be a more reasoned alternative to grading than strict adherence to either an absolute or relative method.

TM025610

standards would best suit the primary function of marking--to provide feedback about academic achievement.' (Kubiszyn & Borich, 1990, p. 144-145). However, there are proponents of the absolute standards grading approach that feel quite strongly about the relative standards approach: 'If a teacher truly graded according to the normal curve and allocated proportions of high and low grades on the basis of the normal curve's properties, such a grading scheme would be truly reprehensible. Procrustean grading proclivities of that sort should definitely be expunged.' (Popham, 1990, p. 371); 'This form of grading, which presupposes that achievement will be normally distributed in a given classroom and that therefore a certain number of A's, B's, C's, D's, and F's must be given in relationship to the normal curve distribution, is a prostitution of statistics (emphasis in the original) and a poor and unfair way to grade.' (Karmel & Karmel, 1978, p. 442). It would appear that knowing-the-principles has not always led to uniform (or even consistent) recommendations for practice.

RELATIVE STANDARDS

To use relative grading methods properly, a teacher needs a larger reference group than the typical classroom. In the usual situation, the mean and standard deviation of scores on a test in this larger group are estimated and used to create standard scores (z-scores) in the classroom sample. Grades are based on these standard scores within the larger population. This is essentially a linear equating of the test scores from the sample (classroom) to the scores of the larger reference group. Classroom sizes are often just not large enough for the teacher to be confident that the class is representative of the population and, thus, standardizing scores within a class is discouraged in most texts.

If a larger reference group is not available (which will be the case when a new test is used with a single class) and there is no written school policy in this regard, it is reasonable to inquire about current distributions of letter grades in the same or similar courses and use this as a guide.

The usual criticism of relative standards, that some students will be predestined to fail, is false when relative standards are correctly utilized.

A necessary supposition for any sort of valid relative evaluation is that consumers of the evaluation be familiar with the reference group. The relative position of an individual within a reference group will be meaningful only when this is the case.

Note that teachers using relative standards also want their tests to be appropriate in difficulty for their students. That is, it is desirable that the proportion of items answered correctly to be reasonable (in the 60-75 percent correct range for most objectively scored achievement tests). Indeed, absolute performance is relevant even when making relative decisions.

ABSOLUTE STANDARDS

The appropriate use of absolute grading standards for a classroom test requires a clearly defined content domain and justified standards of performance (Gronlund & Linn, 1990). There is a large literature on standard-setting procedures, but none, to my knowledge, advocates the use of uniform and seemingly arbitrary standards for all tests and all teachers as is the case in many school district grading policies. The need for a well-defined content domain would seem to limit the content of the test to primarily lower cognitive level items since the item sample on the test needs to be representative of the content domain and higher level skills are neither finite in number nor easily delimited (Hanna & Cashin, 1987; Hanna, 1993). Furthermore, 'A rigid adherence to the conventional percentages could discourage teachers from including many items from the higher taxonomy levels, resulting in an educational disservice to students both in terms of instruction and evaluation.' (Hopkins, Stanley, & Hopkins, p. 323).

A major appeal of absolute grading standards is that the student's performance is measured relative to course content and is thus more meaningful than a normative standard. In addition, all of the students may possibly have high levels of performance and the system is intrinsically less competitive than relative standards. That is, the system appears to be more meaningful, optimistic, and democratic or egalitarian. However, it might be noted that only with this type of grading system do we see classes in which the majority of the students receive 'D's and 'F's.

Consumers of a grade determined by absolute standards are necessarily assumed to be familiar with the larger content domain from which the test items form a sample. It is only then that the portion of the content domain that an individual knows becomes meaningful.

GRADING PRINCIPLES IN PRACTICE

Appropriate use of either absolute or relative standards in practice is extremely difficult and/or limiting. In a school district with grading policies '93% and above is an A, etc.' teachers are forced to test at relatively low cognitive levels with easy test items when current practice emphasizes higher level skills. Ironically, administrators would likely prefer the results of a relative standards approach to grading with the more consistent distributions of A's to F's across teachers.

Using relative standards appropriately, however, requires that a teacher use the same tests with many classes to arrive at the necessary norms, resort to within-class relative performance, or use some form of supplementary information concerning an appropriate distribution of grades. Since instruction is often tailored to the interests and abilities of the students in an individual class, it may be difficult or even unwise to use identical tests. An important consideration in the sole use of relative standards for some is the competitive nature of this

approach to grading.

Practitioners may recognize some of these problems and deliberately choose to compromise between absolute and relative standards. Among the more popular, but inappropriate, methods of compromise are the following:

1. random-gaps: inspection of a score distribution for zero frequencies that are used to define the standard(s) for the test
2. norming-on-the-outlier: raising everyone's score by the difference between a perfect score and the highest obtained score
3. adjusting observed number-correct scores to a desired mean (usually with a linear transformation); note that using standard scores within a class, a common form of relative decision making, simply adjusts the mean and standard deviation to the desired values of 0 and 1 with a linear transformation
4. eliminate very difficult items, 'adjust' partial credit, make the next test easier or more difficult, throw-out the lowest quiz, and so on.

The first two methods are described in more detail in Ebel & Frisbie (1991). There are better ways to arrive at a compromise. For example, a very good case can be made for using relative standards for A-D decisions and absolute standards for the single D-F decision (Terwilliger, 1989; Gronlund & Linn, 1990).

The value of a compromise also is seen in recommendations for formal standard-setting procedures (some of which are applicable to the classroom) in that normative data are called upon to 'inform' the decision or adjust the criterion (Mills & Melican, 1988; Beuk, 1984; De Gruijter, 1985; Hofstee, 1983). In his text on evaluating student achievement, Cangelosi (1990) calls for a compromise of absolute and relative grading standards by setting up grey or buffer zones using a SEM (somewhat similar to the method of Hofstee). Johanson (1992) recommends a compromise that is essentially a variation of Beuk's (1984) method.

Since teachers often have varied instructional objectives, it might make sense to evaluate one unit of work with a test using absolute standards (perhaps the content domain is finite and easily defined) and another unit with a test using relative standards (perhaps the content domain is neither finite nor easily defined). When these letter grades are combined into a summative grade, still another form of compromise becomes apparent.

In general, experts seem to agree that distributions of letter grades should be both reasoned and reasonable. That is, 'Absolute standards should be tempered by the performance of the class as a whole.' (Karmel & Karmel, 1978, p. 445). Others (e.g., Frisbee & Waltman, 1992) note that, regardless of grading method, '... grades from the past few years are probably the best indication of what current outcomes should be like.'

A common implementation of the preceding advice by the more

experienced teacher is to ostensibly use absolute standards but to carefully select tests and test items with a difficulty level that will yield a desired, or at least acceptable, distribution of grades. Now, when a student on such a test is judged to have earned a 'B' (perhaps 85% of the items correct and at the 70th percentile), is it better to give this level of achievement an absolute or relative interpretation? Perhaps it is best to acknowledge that this grade represents a blending of (nominally) absolute and (experience-based) normative standards.

PSYCHOLOGICAL ISSUES: INDIVIDUAL DIFFERENCES AND LEARNING THEORY

How is it that there is such disagreement both between and within the various groups that constitute the educational community regarding appropriate grading standards? Perhaps there are underlying assumptions that predispose an individual towards one method of standard-setting and away from the other.

Are the individual differences that we observe in achievement tests real or merely a by-product of the test? Hanna and Cashin (1987) state: 'If there is anything that psychologists agree upon, it is that individuals differ. This has profound implications for instruction. Effective teaching helps all students develop their talents to the maximum: it increases individual differences (emphasis in the original).' Contrast this statement with the following: '..the expectation that instruction causes a normal distribution of ability is apparently rooted in a belief in the inevitability of cognitive inequality of human beings... Apparently, to make everyone masters of calculus or appreciators of literature would be a great lie.' (Cohen, 1987, p. 19). Clearly, the former position supports the use of relative standards while the later position would imply the use of absolute standards.

Selecting items with large positive discrimination indices (as a norm-referenced test developer is want to do) does, in fact, magnify individual differences. On the other hand, teaching and testing for mastery of a list of vocabulary words may indicate that all of the students in a class have mastered the task. The apparent contradiction would seem to be rooted in the item difficulty and cognitive level involved; with appropriate instruction, we may all be able to reasonably master of certain factual material, but there will likely be individual differences in our higher level skills if these are tested.

Still another relevant factor in the preference for relative or absolute standards might be how we conceive of learning. Shepard (1991, p. 9) found that: '...approximately half of all measurement specialists operate from implicit learning theories that encourage close alignment of tests with curriculum and judicious teaching of tested content.' Her conclusion was that 'These beliefs, associated with criterion-referenced testing, derive from behaviorist learning theory...'. If implicit learning theories and/or beliefs regarding individual differences tend to predispose teachers and others towards the use of

absolute or relative grading standards, then these assumptions should be raised to the level of consciousness and acknowledged.

CONCLUSIONS

We know how to grade, but what should we recommend in practice? Since grading practices will nearly always be forced to depart from the appropriate use of either absolute or relative standards alone, what can we recommend to practitioners? A compromise of absolute and relative standard-setting methods is warranted in that

1. The potential excesses of either method are counterbalanced by the other.
2. The cognitive level of most of our classroom tests tend to (or should) include both factual and higher level skill items.
3. Teacher's learning theories may resemble a patchwork quilt of behaviorism and other cognitive structures.
4. Students may all be masters of some portions of the curriculum at some level, but individual differences will likely prevail in other areas and/or at other levels.
5. We may feel that the typical consumer of a grade has a partial conception of both the content domain and the student population, but a comprehensive knowledge of neither.
6. Compromise may well be most consistent with current practice in that many school districts have absolute standards and many of those teachers feel the need to 'adjust' their scores.
7. Compromise may well enhance the validity of grades in the sense that both the consumers (parents) and the creators (teachers) of grades agree that grades admit both relative and absolute interpretations (Waltman & Frisbie, 1994).
8. Finally, records of grades (transcripts) do not typically indicate how the grade was calculated and thus how to correctly interpret the grade. Further, when they do, (report cards with a legend such as 93% and above is an 'A', etc.), there is reason to believe that these standards are often modified in practice. Consumers of grades may well (and correctly) assume that a letter grade reflects <u>both</u> relative and absolute performance to a reasonable extent. That is, an 'A' typically represents both very good relative and very good absolute performance while a 'D' typically represents <u>both</u> poor relative and poor absolute performance.

This last reason for a compromise is sometimes used as an argument against mediation in that the use of a compromise method of assigning letter grades does alter (and, admittedly, complicate) the interpretation of the resulting grades from that

which would be possible with a purely norm-referenced or criterion-referenced approach (Frisbie & Waltman, 1992; Waltman & Frisbie, 1994). Nonetheless, the limited applicability and potential unfairness of the pure forms of either absolute or relative standards may, as practitioners keep telling us, make compromise methods a necessity.

Brookhart (1991) discusses a conflict between grading practice and the recommendations of measurement specialists with regard to the inclusion or exclusion of 'effort' as a component of a letter grade. In brief, she notes that the uses of grades vary considerably and that this is a contributing factor to the confusion and disagreement. Perhaps this is relevant to the current debate in that some uses of grades may rely on a more normative interpretation (certain selection decisions, perhaps) while other uses (possibly as prerequisites) may be more absolute in nature.

In short, 'The best advice for the teacher is to keep in mind both absolute and normative conceptualizations of mastery...In formal terms, this means reconciling the insights provided by judgments about test content and by judgments about groups.' (Shepard, 1983).

REFERENCES

Airasian, P. W. (1991). Classroom assessment. New York, NY: McGraw-Hill.

American Federation of Teachers, National Council on Measurement in Education, and National Education Association. (1990). Standards for teacher competence in educational assessment of students. Educational Measurement: Issues and Practice, 9(4), 0-32.

Beuk, C. H. (1984). A method for reaching a compromise between absolute and relative standards in examinations. Journal of Educational Measurement, 21, 147-152.

Brookhart, S. M. (1991). Grading practices and validity (letter). Educational Measurement: Issues and Practice, 10(1), 35-36.

Cangelosi, J. S. (1990). Designing tests for evaluating student achievement. New York, NY: Longman.

Cohen, S. A. (1987). Instructional alignment: Searching for a magic bullet. Educational Researcher, 16(8), 16-20.

De Gruijter, D. N. M. (1985). Compromise models for establishing examination standards. Journal of Educational Measurement, 22, 263-269.

Ebel, R. L., & Frisbie, D. A. (1991). Essentials of educational measurement (5th ed.). Englewood Cliffs, NJ: Prentice Hall.

Frisbie, D. A., & Waltman, K. K. (1992). Developing a personal grading plan. Educational Measurement: Issues and Practice, 11(3), 35-42.

Gronlund, N. E., & Linn, R. L. (1990). Measurement and evaluation in teaching (6th ed.). New York, NY: Macmillan.

Hanna, G. S., & Cashin, W. E. (1987). Matching instructional objectives, subject matter, tests, and score interpretations (Idea Paper No. 18), Manhattan, KS: Kansas State University, Center for Faculty Evaluation and Development.

Hills, J. R. (1981). Measurement and evaluation in the classroom (2nd ed.). Columbus, OH: Merrill.

Hofstee, W. K. B. (1983). The case for compromise in educational selection and grading. In S. B. Anderson, & J. S. Helmick (Eds.), On educational testing (pp. 109-127). San Francisco: Jossey-Bass.

Hopkins, K.D., Stanley, J.C.,& Hopkins, B.R. (1990). Educational and psychological measurement and evaluation (7th ed.). Englewood Cliffs, NJ: Prentice-Hall.

Johanson, G. A. (1992, April). A compromise grading model for classroom tests. A paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Karmel L. J., & Karmel, M. O. (1978). Measurement and evaluation in the schools (2nd ed.). New York, NY: Macmillan.

Kubiszyn, T. & Borich, G. (1990). Educational testing and measurement (3rd ed.). Glenview, IL: Scott, Foresman.

Mehrens W. A., & Lehmann, I. J. (1984). Measurement and evaluation in education and psychology (3rd ed.). New York, NY: Holt, Rinehart and Winston.

Mills, C. N., & Melican, G. J. (1988). Estimating and adjusting cutoff scores: Features of selected methods. Applied Psychological Measurement, 1(3), 261-275.

Oosterhof, A. C. (1990). Classroom applications of educational measurement. Columbus, OH: Merrill.

Popham, W. J. (1990). Modern educational measurement: A practitioner's perspective (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Sax, G. (1989). Principles of educational and psychological measurement and evaluation (3rd ed.). Belmont, CA: Wadsworth.

Shepard, L. A. (1991). Psychometricians' beliefs about learning. Educational Researcher, 20(6), 2-16.

Shepard, L. A. (1983). Standards for certification and placement. In S. B. Anderson, & J. S. Helmick (Eds.), On educational testing (pp. 61-90). San Francisco: Jossey-Bass.

Terwilliger, F. S. (1989). Classroom standard setting and grading practice. Educational Measurement: Issues and Practice, 8, 15-19.

Waltman, K. K., & Frisbie, D. A. (1994). Parent's understanding of their children's report card grades. Applied Measurement in Education, 7(3), 223-240.